DOCUMENT RESUME

ED 179 600 TM 009 991

AUTHOR Echternacht, Gary
TITLE The Norm Referenced Evaluation Model and the Regression Effect.
PUB DATE Apr 79
NOTE 19p.; Paper presented at the Annual Meeting of the American Educational Research Association (63rd, San Francisco, CA, April 8-12, 1979)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Achievement Gains; Comparative Statistics; Educational Assessment; Elementary Secondary Education; *Evaluation Methods; *Mathematical Models; National Norms; Norm Referenced Tests; Post Testing; Pretesting; Standard Error of Measurement; Statistical Analysis; *Statistical Bias
IDENTIFIERS Elementary Secondary Education Act Title I; *Regression (Statistical); *RMC Models

ABSTRACT

The role that measurement error plays in the regression effect is discussed with particular reference to the RMC models for evaluating the effects of Elementary Secondary Education Act Title I programs. The norm referenced evaluation model assumes a theory of growth where the relative ranking of students remains the same from pretest to posttest in the absence of any supplementary treatment (the equipercentile growth assumption). However, if the pretest is used to select low scoring candidates for Title I programs, there is certain to be a preponderance of negative errors among these scores, so that any unbiased posttest is certain to indicate growth even in the absence of any treatment. However, it is recommended that Title I programs use three tests: a selection test, an independent pretest for the selected group, and a posttest. A regression problem remains, however, if the correlation between the selection test and the pretest is not the same as the correlation between the selection test and the posttest. (Author/CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED179600

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

The Norm Referenced Evaluation Model and the Regression Effect

Gary Echternacht

Educational Testing Service

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

G. Echternacht

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM009 991

Presented at the Annual AERA Meeting 1979

San Francisco, CA

Abstract

As required by P.L. 95-158, school districts are required to use one of three evaluation models to evaluate their ESEA Title I programs. The most frequently used model is the norm referenced model, in which program impact is indicated in terms of gain in normative standing from pretest to posttest. To overcome the regression effect found when using gain scores, it has been suggested that by having separate selection and pretest administrations, the regression effect might be either eliminated completely or reduced to an inconsequential amount. This paper disputes that claim by providing an exact definition of the regression effect in the three variable case and by showing that the assumptions used to conclude no regression are inconsistent with the definition. Examples are given of ways in which the amount of regression can be determined from knowledge of the correlational structure of the variables.

The Norm Referenced Evaluation Model And The Regression Effect

As required by P.L. 95-158, school districts receiving ESEA Title I funds to conduct reading, language, and mathematics projects in grades 2 through 12 are required to evaluate their projects by using one of three evaluation models. Each model attempts to estimate the extent to which student posttest performance exceeds the level of performance one would expect had there been no Title I project. Two of the models estimate this "no treatment performance" by statistics obtained from comparison groups (see Tallmadge and Wood, Note 2); in the third model estimates of no-treatment performance are derived from the norms of nationally standardized achievement tests (see U.S. Department of Health, Education, and Welfare, Office of Education, Note 1). It is called the norm referenced model, where a treatment effect is defined in terms of student gain in percentile rank.

The norm referenced evaluation model assumes a theory of growth where the relative ranking of students remains the same from pretest to posttest in the absence of any supplementary treatment. This is often referred to as the equipercentile growth assumption (Tallmadge and Wood, Note 2). For example, when using the model one asserts that a project would have had no effect if both the pretest and posttest score averages for the treated group correspond to the same percentile rank. Title I projects, as well as other categorically aided compensatory education projects, are intended to be supplementary, having the goal of raising the academic achievement levels of students beyond the level normally occurring. With the norm

referenced model, demonstrating a gain in the normative standing of the Title I group is considered evidence of a positive project effect.[1]

When using the norm referenced model, normalized standard score scales are used for reporting pretest and posttest scores.[2] If x represents the pretest score and y the posttest score, and $\bar{X}$, $\bar{Y}$, S_x, and S_y represent the norms groups means and standard deviations, then in the norm referenced model there would be no effect when

$$\frac{x - \bar{X}}{S_x} = \frac{y - \bar{Y}}{S_y}$$

If we knew x, we could solve for $y = y_o$ and obtain $y_o = \bar{Y} + (S_y/S_x)(x - \bar{X})$, which is the no-treatment expectation under the model. In the federal reporting system $\bar{X}$ and $\bar{Y}$ are both set equal to 50, and S_x and S_y are set equal to 21.06. Without loss of generality, in this paper we will assume that $\bar{X} = \bar{Y} = 0$ and $S_x = S_y = 1$. In such an instance $y_o = x$, and our treatment effect becomes the simple gain $y - x$.

---

1 The Users Guide (Note 2) suggests that more valid evaluation models, such as those suggested by Rubin (1977), be considered and rejected before using the norm referenced model. Nonetheless, because of the simplicity of the norm referenced model, and because of its similarity to current practice, most evaluators reject the more valid models.

2 Normalized standard score scales are used because the underlying trait is believed to be normally distributed. By normalizing the standard scores it is believed that the result will be an equal interval scale.

The regression effect and measurement error

In discussing the pitfalls of using simple gain scores, many authors (e.g., Lord, 1967) have noted the regression effect. Although the concept is often presented intuitively or pictorally, in educational evaluation the regression effect indicates the extent to which a student's test score might be expected to increase or decrease on subsequent retesting, when a standard score scale is used. The regression effect is due to the lack of perfect correlation between x and y, or more exactly to the difference between initial standing (x) and an estimate of posttest standing that is obtained from a regression equation. We can express this difference as

$$\begin{aligned} \text{regression effect at } x_o &= R(x_o) \\ &= E(y|x_o) - x_o \\ &= r_{xy}x_o - x_o \\ &= (r_{xy} - 1)x_o \end{aligned} \qquad (1)$$

The symbol $E(y|x_o)$ indicates the conditional expectation of y given the observation that $x = x_o$ and is more commonly thought of as the regression function of y on x. Since $\bar{X} = \bar{Y} = 0$ and $S_x = S_y = 1$, we have the simple equation $E(y|x_o) = r_{xy}x_o$. The difference between the conditional expectation or predicted posttest and the observed pretest standing is the amoung of regression we expect at x_o.

When x and y are achievement tests, there is often concern over the role that measurement error plays in determining the extent of the regression

effect. If we make the usual assumptions of classical test theory (e.g. Gulliksen, 1950) where observed scores are separated into true and error components, with the errors uncorrelated with either true scores or other errors, then $x = t_x + e_x$, $y = t_y + e_y$. and if we look at the expected regression effect, where the expectation is taken over repeated measurements, we have

$$\begin{aligned} E\text{ (regression effect at } x\text{)} &= E(R(x)) \\ &= E(E(y|x) - x) \\ &= E((r_{xy} - 1)x) \\ &= (r_{xy} - 1)t_x \end{aligned}$$

If we denote the reliability of the tests by $S^2_{t_x} = r_{xx}$ and $S^2_{t_y} = r_{yy}$, then

$$\begin{aligned} E(R(x)) &= (r^{\frac{1}{2}}_{yy} r^{\frac{1}{2}}_{xx} r_{t_x t_y} - 1)t_x \\ &= (r^{\frac{1}{2}}_{yy} r^{\frac{1}{2}}_{xx} - 1)r_{t_x t_y} t_x + (r_{t_x t_y} - 1)t_x \end{aligned}$$

The right most term of the above equation gives the amount of expected regression that would have occurred had we obsered the true scores t_x and t_y rather than x and y . The left most term of the right side of the equation indicates the portion of the regression effect that might be attributed to measurement error in x and y . It is possible then to compute the percentage of the regression effect that we might attribute to measurement error by

$$\left[\begin{array}{c}\% \text{ of regression attributed} \\ \text{to measurement error}\end{array}\right] = ((r^{\frac{1}{2}}_{yy} r^{\frac{1}{2}}_{xx} - 1)r_{t_x t_x} / (r^{\frac{1}{2}}_{yy} r^{\frac{1}{2}}_{xx} r_{t_x t_y} - 1))100\% \quad (2)$$

If $r_{t_x t_y} = 1$, then we would attribute all of the regression effect to measurement error, or we would conclude that the imperfect correlation between x and y was due entirely to measurement error.

In cases where $r_{t_x t_y}$ is close to one, some might argue that the regression effect would be reduced substantially (or eliminated completely if $r_{t_x t_y} = 1$) by requiring that the pretest administration not be used for selecting students for the project. Title I regulations require that only the most needy students be placed in the project, which usually means that those scoring below some predetermined cut-off on a standardized test are selected. If the test administration used for selection were also used as the pretest measure, we would expect project gains when using the norm referenced model to be inflated estimates of treatment effect since the cut-score is usually less than $\bar{X}$ - i.e., we have $R(x)>0$ for all $x < 0$. Advocates of the norm referenced model argue that if x_1 represents the selection test, x_2 the pretest, and y the posttest, and if true scores are designated by t_1, t_2, and t_y respectively, and if $r_{t_2 t_y} = 1$ and the reliabilities $r_{11} = r_{22} = r_{yy} = r_{**}$, then the correlation between selection test score and gain score becomes

$$r_{x_1(y - x_2)} = \frac{r_{**}(r_{t_1 t_y} - r_{t_1 t_2})}{\sqrt{2(1 - r_{**} r_{t_2 t_y})}}$$

Since $r_{t_2 t_y} = 1$ and $r_{11} = r_{22} = r_{yy}$, we have that $t_1 = t_2 = t_y = t$ and that $r_{t_1 t_y} = r_{t_1 t_2} = r_{t_2 t_y} = 1$ so that the correlation between the selection test and gain score is zero. One might then contend that the regression effect has been eliminated by calculating gains based on a test

administration other than the selection test Even if r_{t_2y} is not exactly one, it might be argued that by giving a separate pretest an inconsequential regression effect will result.

The regression effect with three tests, x_1, x_2 and y.

But does the above situation with $r_{t_2t_y} = 1$ and $r_{11} = r_{22} = r_{yy} = r_{**}$ mean that the regression effect has been eliminated by having separate selection test and pretest administrations? To adequately address this question we must first generalize our definition of the regression effect from the case where we have only a pretest and a posttest to the case where we have selection test, pretest, and posttest. We now define the regression effect by generalizing equation (1) to

$$\begin{aligned} \text{regression effect at } x_2 &= R(x_2) \\ &= E(y \mid x_1, x_2) - x_2 \\ &= b_1 x_1 + b_2 x_2 - x_2 \\ &= b_1 x_1 + (b_2 - 1)x_2 \end{aligned} \tag{3}$$

where b_1 and b_2 are the regression coefficients for x_1 and x_2 in the regression of y on x_1 and x_2. The above definition is an exact statement of the intent of having separate test administrations for selection and pretesting. In the context we have established, it says that the expected posttest score is exactly the score that we observed at pretest, just as one assumes in the norm referenced model. Our expectation is based on knowledge of the selection test and pretest scores that are observed.

The coefficients in equation 3 can be written in terms of correlations by $b_i = (r_{iy} - r_{jy}r_{12})/(1 - r_{12}^2)$. If we represent each pretest score as a function of the selection test score by $r_{12}x_1 + h$ where h is the residual from the conditional expected value of x_2 (or regression function), we can simplify equation (3)

$$\begin{aligned} R(x_2) &= b_1x_1 + (b_2 - 1)(r_{12}x_1 + h) \\ &= (b_1 + r_{12}b_2 - r_{12})\, x_1 + (b_2 - 1)h \\ &= (r_{1y} - r_{12})x_1 + (b_2 - 1)h \end{aligned} \qquad (4)$$

Earlier we had determined that when $r_{t_2y} = 1$, $r_{11} = r_{22} = r_{yy} = r_{**}$, we have $r_{x_1(y - x_2)} = 0$, leading us to expect that $R(x_2)$ would be zero. If we know that $t_1 = t_2 = t_y = t$, using equation (3) to find the expected regression effect (the expectation taken over repeated trials), we have

$$\begin{aligned} E(R(x_2)) &= (b_1 + b_2 - 1)t \\ &= \frac{(r_{1y} + r_{2y} - r_{12} - 1)t}{1 + r_{12}} \\ &= \frac{(r_{**} - 1)\, t}{(r_{**} + 1)} \end{aligned} \qquad (5)$$

which is equal to zero if either $t=0$ or $r_{**}=1$. That is, if either our true score is at the mean for all tests or if the tests are perfectly reliable, then there will be no regression effect in the norm referenced model. If our tests are perfectly reliable, however, we will be able to observe that the true scores, or in this case the observations, do not vary over the testings.

The relationship to the definition of the regression effect given by (3) to that of $r_{x_1(y-x_2)}$ may be seen by examining equation (4). If $r_{x_1(y-x_2)} = 0$, then r_{1y} must equal r_{12}. This would mean that the coefficient of x_1 in (4) would be zero, but the coefficient of h, the residual from the regression of x_2 on x_1 would be $(r_{2y} - 1)/(1 - r_{12}^2)$, which would be zero only if $r_{2y} = 1$ and $r_{12} \neq 1$. Using our above definition of the regression effect, $R(x_2)$ would be zero if $h=0$. But this would mean that the relationship between pretest and selection test were perfect in the sense of having a correlation of one, or that x_2 was not an observed value at all, but rather the conditional expectation, or the value of the regression function for the regression of x_2 on x_1. In the norm referenced model then, we could eliminate the regression effect by using $E(x_2 | x_1) = r_{12}x_1$ as pretest scores, rather using the observed x_2 as pretest scores.

If we return to equation (4), we see that the regression effect will be eliminated by requiring that the selection test be separate from the pretest in a number of special cases. If the observed selection test and pretest scores are both zero, then there is no regression effect. If $b_1 = 0$ and $b_2 = 1$, there will be no regression effect. The coefficient b_1 being equal to zero is equivalent to saying that $r_{1y} - r_{2y}r_{12} = 0$, or that the partial correlation $r_{1y.2} = 0$. The coefficient b_2 will be unity when $r_{2y} - r_{1y}r_{12} = 1 - r_{12}^2$ or when the partial correlation $r_{2y.1} = ((1 - r_{12}^2)/(1 - r_{1y}^2)$. It is important to note that both of these conditions must be met to have no regression effect and that in general such conditions will not hold. The regression effect also will be

zero when

$$x_2 = \frac{(r_{1y} - r_{2y}r_{12})x_1}{1 - r_{12}^2 - r_{2y} + r_{1y}r_{12}}$$

or for special combinations of x_1 and x_2 that are determined by the correlational structure of the variables. Again, this is a particular case and cannot be expected in general.

Examples of the extent of the regression effect in the norm referenced model.

Most work on the regression effect and the norm referenced evaluation model has appeared in unpublished memos and papers and has been conducted under government contract.[1]

Two examples are given in Table 1 that illustrate the effects of measurement error on the size of the regression effect. Both examples use correlations obtained from test publisher longitudinal normings. The correlations given are those between scores obtained in fall and spring administration of the publishers tests. Swinton (Note 3) has presented data for the grade 2 Circus C and D instruments in reading. The reliabilities presented in the table for Swinton's data are the KR_{20} values.

---

[1]There has been work at the theoretical level, computer simulation level, and analyzing empirical data level. It is not feasible to list all those who have contributed although it should be said that most of the work has been performed by RMC Corporation, who were responsible for the development of the Title I evaluation models, staff at the Title I regional Technical Assistance Centers, and members of RMC's technical advisory committee.

Roberts (Note 4) has obtained similar statistics for scores in Total Reading on level 16 of the California Achievement test, which is given for grade 6. Roberts obtained alternate forms reliabilities, however, which he further corrected for some range restrictions. The percent of regression that might be attributed to measurement error was determined for each set of data by using equation (2). Since the norm referenced model is most frequently used for Title I evaluations, the amount of regression that we might expect had we observed true scores is expressed in Normal Curve Equivalent (NCE) scores, assuming that a pretest score of 20 was observed. The NCE is a normalized standard score with mean 50, standard deviation 21.06. Title I evaluations report project impacts in terms of NCE gains.

Insert Table 1
about here

The two reported samples differ in their conclusions. From Swinton's data, we conclude that most of the regression effect would remain had we observed true scores. In this case we suspect that by using a separate testing for selection and pretest the regression effect will be reduced only marginally. Roberts' data, which is consistent with Swinton's at lower grade levels, provides us with another conclusion. The estimated true score correlation is nearly one, and assuming an initial NCE of 20, only about .3 NCE units of regression might be expected had true scores been observed. If we assume that $r_{t_x t_y} = 1$, selection, pretest, and posttest reliabilities are equal, and use equation (5), however, to indicate how much regression we might expect when using a separate test for

selection and pretest, we conclude that at a true score of 20 we might expect a regression effect of about 1.3 NCEs, about 4 times larger than we might conclude from using equation (2).

If we knew the correlational structure of the variables, we could determine how much of the regression effect a separate pretest was likely to remove. Longitudinal data, such as that given by Hilton and Berglund (1974), suggest that correlations between test scores deteriorate as the time between measurement of the variables increases.

If we assume that $r_{12} = .5$, $r_{1y} = .4$, and $r_{2y} = .6$, we can calculate the extent of the regression effect when using the norm referenced evaluation model with separate selection tests and pretests. An indication of the size of the regression effect is provided in Table 2, where equation (3) is computed for various combinations of selection test and pretest scores. The dashed line in the body of the table indicates combinations of x_1 and x_2 where there is no regression effect. To translate entries of the table into NCEs, we need only to multiply by 21.06. Thus, in the example, if $x_1 = -1.2$ (NCE = 25) and $x_2 = -.75$ (NCE = 34) we would expect a regression effect of about $.23 \times 21.06 = 4.8$ NCEs. Although the norm referenced model would suggest that the posttest NCE should be the same as the pretest NCE if there were no special intervention--i.e., we expect the posttest NCE score to be 34--our definition of the regression effect suggests that the expected posttest should be 38.8 rather than 34.

Insert Table 2
about here

Summary

In this paper we have attempted to describe the role that measurement error plays in the regression effect and one way that has been suggested for overcoming the regression effect in the norm referenced evaluation model. This requires using as a pretest a test administration that was not used for selection.

The rationale given for this is that if pretest and posttest true scores correlate perfectly and if the pretest and posttest reliabilities are the same, then the correlation between selection test and the gain score will be zero. Even if the correlation is not exactly unity, the regression effect should be negligible if the true scores correlate high.

To examine the validity of this claim, we have proposed a definition of the regression effect for the norm referenced evaluation model when selection test and pretest are different. This definition is consistent with the assumptions of the norm referenced model and is based on the conditional expectation of the posttest scores given the selection test and pretest scores, or the regression of posttest on pretest and selection test. Using this definition, we find that $r_{x_1(y-x_2)} = 0$ is equivalent to saying that there is no regression effect only when x_2 is the conditional expectation of x_1. In general, there are only specific combinations of selection test and pretest scores at which the regression effect will be zero. The magnitude and the extent of the regression will depend on the correlational structure of the variables, and in most cases the regression effect will be larger than we would expect by examining only pretest-posttest correlations and their reliabilities.

Reference Notes

1. U.S. Department of Health, Education, and Welfare, Office of Education A practical guide to measuring project impact on student achievement. Washington, D.C.: U.S. Government Printing Office, 1975.

2. Tallmadge, G. K., & Wood, C. T. User's Guide: ESEA Title I evaluation and reporting system. Mountain View, CA,: RMC Research Corporation.

3. Swinton, S. Some good news and some bad news for the Model A regression assumption. Unpublished memorandum, 1979.

4. Roberts, O. Unpublished tables from CTB/McGraw-Hill CAT Fall 1976 - Spring 1977 standardization, pre-post correlations, alternate forms reliabilities, reliability corrected for range, and correlation corrected for attenuation, 1979.

References

Hilton, T and Berglund, G. Sex differences in mathematics achievement - a longitudial study. Journal of Education Research. 1974, 67(5), p. 231-237.

Hodges, W. The Worth of the Follow Through experience. Harvard Educational Review, 1978, 48(2), p. 186-191.

Lord, F. Elementary models for measuring change. In Harris, C.(ed), Problems in Measuring Change. Madison: U. of Wisconsin Press, 1967. p. 21-38.

Rubin, D. Assignment to treatment group on the basis of a covariate. Journal of Educational Statistics. 1977, 2(1), p. 1-26.

Table 1

Illustration of the effects of measurement error on regression to the mean in two samples

Characteristic	Swinton Reading (Circus) C and D	Roberts Total Reading (CAT) C and D
Grade level	2	6
r_{xy}	.72	.88
r_{xx}	.94	.89
r_{yy}	.93	.89
$r_{t_x t_y}$	.77	.99
% of regression attributed to measurement error	18%	92%
Regression to mean in NCES, x = 20	6.9	.3

Table 2

Computation of the regression effect for various levels of selection test and pretest scores where $r_{1y} = .4$, $r_{2y} = .6$, and $r_{12} = .5$*

Selection Test x_1	% tile of x_1	Pretest Score x_2								
		.3	0	-.15	-.3	-.45	-.6	-.75	-.9	-1.2
0	50	-.14	0	.07	.14	.21	.28	.35	.42	.56
-.3	38	-.18	-.04	.03	.10	.17	.24	.31	.38	.52
-.6	27	-.22	-.08	-.01	.06	.13	.20	.27	.34	.49
-.9	18	-.26	-.12	-.05	.02	.09	.16	.23	.31	.45
-1.2	12	-.30	-.16	-.09	-.02	.05	.12	.20	.27	.41
-1.5	7	-.34	-.20	-.12	-.05	.02	.09	.15	.23	.37
-1.8	4	-.38	-.23	-.16	-.09	-.02	.05	.12	.18	.33

*$E(y \mid x_1, x_2) = .13x_1 + .53x_2$